MEASURING THE INTERACTION BETWEEN BASSOON AND HORN PLAYERS IN ACHIEVING TIMBRE BLEND

Sven-Amin Lembke **Scott Levine** **Martha de Francisco** **Stephen McAdams**
Centre for Interdisciplinary Research in Music Media and Technology, Schulich School of Music, McGill University
sven-amin.lembke@mail.mcgill.ca

ABSTRACT

Our study investigates the interactive relationship between bassoon and horn players in achieving timbre blend during musical performance. The interaction is studied in a behavioral experiment, measuring the timbral adjustments performers employ. Several timbre descriptors serve as acoustic measures, quantifying global and formant-based spectral-envelope properties. Furthermore, musicians' self-assessment of their performances is measured through behavioral ratings. The performances are investigated across four factors, i.e., room acoustics, communication directivity, musical voicing, and leading vs. accompanying roles. Findings from ANOVAs suggest that differences in role assignments and communication directivity between performers lead to timbral adjustments. These effects are more pronounced for horn than for bassoon and performer interdependencies appear to be most important for unison voicing.

1. INTRODUCTION

In orchestration practice, composers rely on their experience and intuition to obtain instrument combinations that lead to blended timbres, i.e., combinations exhibiting higher degrees of perceptual fusion. Previous research on timbre blending has emphasized explanations of the degree of blend through correlations with acoustic instrument properties. However, the contribution of musical performance factors to the actual realization of timbre blend remains largely unexplored. Past investigations of timbre blending between orchestral instruments have instead primarily employed stimuli that were created by a mix of solo-instrument recordings [1, 2], with their findings not fully extending to more realistic scenarios. In musical practice, blend is always performed by two or more musicians in an interactive relationship that allows for timbral adjustments between performers. Our investigation focuses on this interactive relationship between two performers attempting to blend together.

A previous investigation of performer interaction focused on synchrony between two pianists [3]. Experimental factors such as performer role or acoustical feedback were investigated, showing asymmetric dependency of players acting as *followers* on the *leading* pianists. Furthermore, under impaired acoustical feedback, performers increasingly relied on visual cues to maintain synchrony, which argues for investigations of performance-related factors involving auditory properties alone to exclude the possibility of visual communication between performers. With regard to common examples from the orchestral repertoire, musicians performing in a blended pairing may involve either doubled performances in (pitch) unison or paired phrases in non-unison. In both scenarios, one of the performers would usually assume the leading role, with that role commonly also being associated with the top voice in non-unison cases. It therefore may be hypothesized that followers would adjust their timbres to the leading performer and not vice versa. Moreover, a general validity of this unilateral dependency should not result in the leader performing differently, if they were to receive no auditory feedback from the follower, as might occur in unfavorable studio or live-performance situations.

Performer interaction in achieving timbre blend is investigated in a behavioral experiment for an instrument combination that finds widespread use in the orchestral repertoire, namely, the combination of bassoon and (French) horn. Orchestration treatises discuss these two instruments as forming a common blended pairing [4–7], with these observations reflected in findings of high degrees of blend in perceptual investigations [1, 2]. The horn is often considered an unofficial member of the woodwind section, bearing a timbral versatility that succeeds in blending with woodwinds, brasses, and even strings. Given the relevance to orchestration practice, the investigation of musical performance situates musicians in approximation to the ecologically valid setting of a concert hall, realized through controlled and reproducible virtual performance environments. The measurement of musical performance is conducted in both behavioral and acoustic domains.

2. METHODS

2.1 Experimental design

The behavioral experiment addresses a series of research questions. The principal aim investigates what instrument-specific adjustments are employed in achieving timbre blend and how these interact in a performance scenario with two musicians. These interactions are furthermore studied as a function of musical and acoustical factors. The experiment is based on a mixed-model design, with the two instruments implemented as a between-participants factor. All remaining factors employ a repeated-measures design, to rule out the possibility that individual differences for instruments and playing technique or style are confounded with the investigated effects for musical and acoustical factors.

2.1.1 Musical factors

Two within-participant, independent variables involve the performer role and the influence of different musical voice contexts. The former considers one instrumentalist taking on the role of *leader*, whereas the other performer acts as *follower*, i.e., takes on an accompanying role. According to the 'voice' factor, musicians either perform a melodic phrase in *unison* or a musically related, two-voice phrase in *non-unison*. The musical excerpts are taken from Mendelssohn-Bartholdy's *A Midsummer Night's Dream*, Op. 61, No. 7 (measures 1-16). In this orchestral excerpt, the chosen instrument combination is featured prominently, with a horn solo being accompanied by two bassoons. All phrases were transposed by a fifth down to A major from the original key of E major, to reduce the impact of player fatigue through repeated performances in high instrument registers. The solo melody functions as the unison excerpt, denoted A; the two accompanying voices serve as the top and bottom voices in the non-unison condition, denoted B and C, respectively, with B being assigned to the leader.

2.1.2 Acoustical factors

Another pair of within-participant variables considers effects for communication directivity between performers and the room-acoustical properties of performance venues. The 'communication' factor assesses the influence of whether both performers are able to hear each other or only the follower hears the leader, denoted *two-way* or *one-way*, respectively. For the 'room' factor, the influence of room acoustics is assessed for two different performance spaces: musicians are simulated as performing in either a large, multipurpose performance space (Music Multimedia Room) or in a mid-sized recital hall (Tanna Schulich Hall). [1]

2.1.3 Procedure

Two participants were tested in a single experimental session, being instructed to perform together to achieve the highest possible degree of blend. Each musician underwent three repetitions of 16 different experimental conditions (four factors by two treatment levels, 2^4), leading to a total of 48 experimental trials. The total duration of the experiment was around two hours, including a break scheduled after half of the trials. To avoid disorientation of musicians through strongly varying performer-role and voice assignments, the musical factors were blocked. Participants assumed the role of either leader or follower throughout the first or second half of the experiment. Furthermore, shorter eight-trial blocks grouped conditions based on voice assignment (e.g., four unison trials, another four non-unison), with the repetitions occurring after each block. For instance, a given participant would begin as leader for 24 trials, performing the first repetition of four unison trials, then proceed to four non-unison trials, followed by the second repetition of the same four unison trials, etc. The four possible block-ordering schemes were counterbalanced across all participants and instruments. The acoustical-factor combinations were encapsulated inside sub-blocks of four trials and randomized in order. Three practice trials were conducted under the guidance of the two experimenters, presenting the experimental conditions encountered at the beginning of individual block-ordering schemes.

[1] Both venues are located at the Schulich School of Music, McGill University. More details under http://www.mcgill.ca/music/about-us/facilities. (Last accessed on March 20, 2013.)

2.1.4 Participants

Sixteen musicians participated in the experiment and were primarily recruited from the Schulich School of Music at McGill University and the music faculty of the Université de Montréal. The bassoonists, three female and five male, had a median age of 21 years (range 18-31). The hornists, six female and two male, had a median age of 20 years (range 17-44). Across both instruments, 10 participants considered themselves as professional musicians, and overall, the musicians reported to play or practice their respective instruments for the median duration of 21 hours per week. All musicians were remunerated with 35 CAD for their participation.

2.1.5 Performance measures

The musical performances were evaluated with the help of a set of behavioral and acoustic measures, which focus on capturing features related to timbre blending. Behavioral measures comprise two ratings that participants provided after each experimental trial. The first rating assessed how well musicians thought they performed individually given their assigned role, on a continuous scale with the verbal anchors *very badly* and *very well*. The second measure acquired ratings on the perceived degree of achieved timbre blend with the other performer, on a continuous scale with the verbal anchors *low blend* and *high blend*. The acoustic measures consist of a number of spectral-envelope descriptors, which are discussed in Section 2.3.

2.2 Technical realization

The experiment was conducted in two research laboratories at the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) at McGill University. Separate laboratory spaces were called for in order to create individual acoustical environments for each participant, ensuring the capture of separate source signals as well as preventing visual cues between performers. Each performance laboratory was treated to be relatively non-reverberant, with a $RT_{60} < 0.5$ s. Performers received instructions to prepare for performances of assigned roles and excerpts and also provided their behavioral ratings through dedicated computer interfaces. Furthermore, the performances were synchronized by attending to a video monitor transmitting a silent conductor cue track.

Each musician's performance was captured through an omnidirectional high-voltage microphone, which were matched across laboratories. Both microphone signals were routed to a control room, where preamplification gain was digitally matched across both performance spaces. The analog signals were converted to 96 kHz / 24-bit PCM digital data, recorded at full resolution for later acoustical analysis and at the same time fed into separate convolution engines, processing the source signals with different sets of binaural impulse responses (IRs). Individualized binaural signals, based on the acoustical factors, were then fed to headphones for each performer. Headphone amplifier volume was held constant, as were the circumaural closed-ear headphones. The convolution introduced a system latency of 805 samples, resulting in delayed arrival of the room feedback by about 8.4 ms, affecting both performers equally and thus not assumed to influence their interaction. The IRs had been previously collected in the concert halls discussed in Section 2.1.2, with RT_{60} for the smaller and larger halls being 1.3 and 2.1 s, respectively. IRs were measured with a binaural head-and-torso system, positioning the excitation source and receiver appropriately for a typical orchestral setup: horns on the conductor's left front side and bassoons on the conductor's right front.

2.3 Acoustic descriptors

For the instruments bassoon and horn, the existence of largely pitch-invariant, local spectral maxima has been reported [8–10], which are also termed *formants* by analogy with the human voice. Furthermore, frequency alignment of formants between instruments has been argued to contribute to the perception of blend [2], with certain aspects of this hypothesis having been replicated in further perceptual investigations [11], confirming the significant contribution of the most prominent formants. On the other hand, global spectral-envelope descriptors, such as the spectral centroid, have also been reported to correlate with the perception of blend [1].

Time-variant spectral envelopes are obtained through *True Envelope* (TE) estimation [12]. The TE algorithm applies iterative cepstral smoothing on STFT-magnitude spectra, with the computed estimates using a constant cepstral order oriented at fundamental frequencies $f_0 \leq 300$ Hz. A formant-analysis algorithm, based on the detection of local spectral maxima and plateaus, i.e., regions of spectral-envelope slopes approximating zero, identifies and classifies up to three formants within a dynamic range of 50 dB. The frequencies of formant maximums (e.g., F_1) serve as descriptors. In addition, the most prominent formant F_1, also termed *main* formant, involves pairs of descriptor frequencies delimiting upper or lower bounds at which the magnitude has decreased by 3 dB or 6 dB (e.g., upper $F_{\overrightarrow{3dB}}$ and lower $F_{\overleftarrow{3dB}}$ bounds relative to F_1). These formant descriptors are illustrated for a spectral-envelope estimate of a single participant's performance in Fig. 1, based on median magnitudes over time. In addition, relative magnitude differences between spectral-envelope regions are considered: for example, $\Delta L_{1vsRest}$ quantifies the level difference between F_1 and the averaged magnitude for frequencies $f > F_{\overrightarrow{6dB}}$. The spectral-envelope estimates furthermore serve as the basis for the computation of the spectral centroid S_c (amplitude-weighted frequency average) and slope S_s (linear regression of the spectrum) [13]. These serve as global, formant-independent descriptors of general spectral trends in the frequency and magnitude dimensions, respectively.

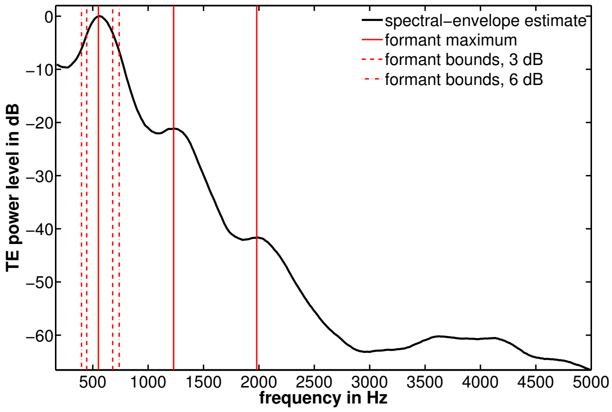

Figure 1. Time-averaged spectral-envelope estimate and its formant description for a single bassoon performance of the unison excerpt.

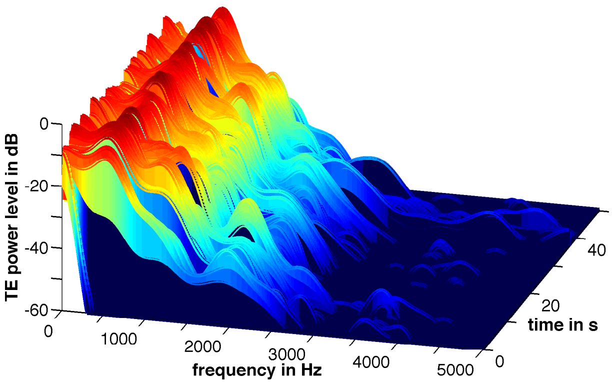

Figure 2. Temporal evolution of *True Envelopes* for the same performance as in Fig. 1.

From qualitative evaluations of spectro-temporal representations for both instruments conducted prior to running the experiment, the chosen spectral-envelope description could be confirmed as capturing relevant features associated with timbral modifications. The main formants F_1 for both instruments are located around 500 Hz and, as illustrated in Fig. 2 for the bassoon, they remain relatively stable across pitch and dynamic range. It also became apparent that the players' control over instrumental timbre is constrained, more so for bassoon than for horn. The main formants of horns are broader, less defined, and more variable in location, which affords horn players greater timbral control. For both instruments, the strongest variability is achieved for changes in dynamic markings, which in the chosen excerpt are limited to a single, notated change (e.g., *crescendo-descrescendo*) in measures 13-14.

3. RESULTS AND DISCUSSION

The strongest trends for effects between instruments and the remaining factors should already become apparent from inferential statistics computed on the behavioral and time-averaged acoustic measures. Moreover, it will not be possible to address more complex effects found across the

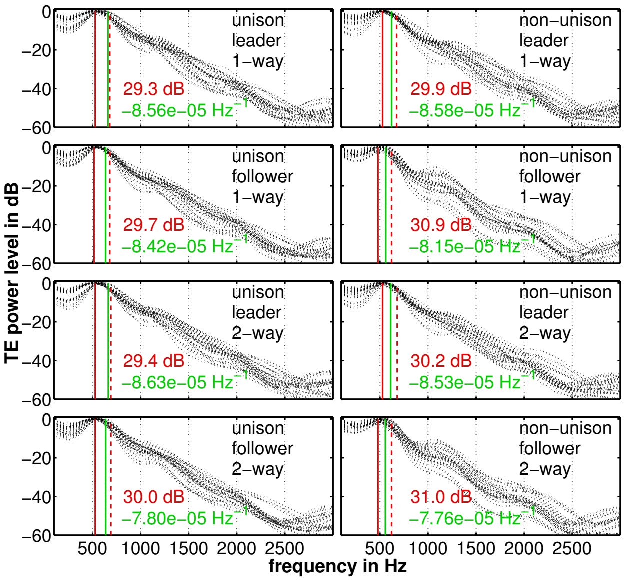

Figure 3. Bassoonists' spectral envelopes and median acoustic-descriptor values per factor combination 'voice' $\times$ 'role' $\times$ 'communication'. Formant description in red: F_1 (solid line), $F_{3dB}^{\rightarrow}$ (dashed line), and $\Delta L_{1vsRest}$ (numerical value). Global descriptors in green: S_c (line) and S_s (numerical value).

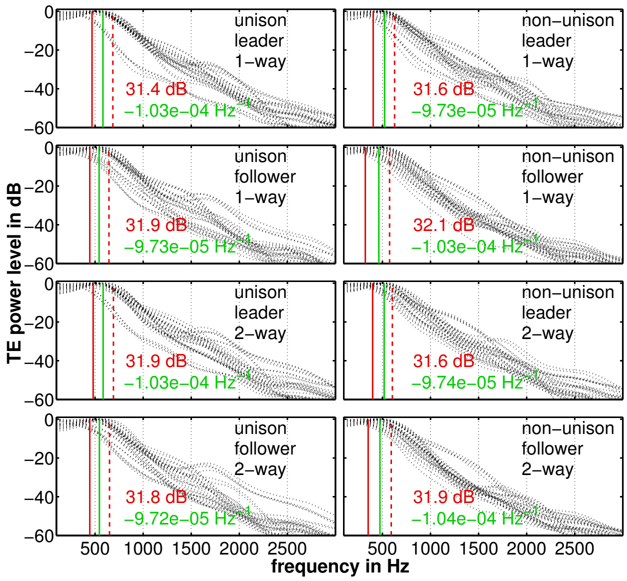

Figure 4. Hornists' spectral envelopes and median acoustic-descriptor values per factor combination 'voice' $\times$ 'role' $\times$ 'communication'. See caption of Fig. 3 for legend.

time course of performances within the scope of this paper. Given that amongst the acquired data some performances were qualitatively better than others, the entire dataset with three repetitions per condition is reduced by retaining only the two performances per participant that yield the highest self-assessed performance ratings. [2] Separate performances are considered as independent cases, i.e., corresponding to a total number of 16 cases (eight performers $\times$ two repetitions) per instrument. Mixed-model ANOVAs involving the between-participants factor 'instrument' and the within-participants factors 'role', 'voice', 'room', and 'communication' were computed, assuming a significance level of $\alpha = .05$. Both behavioral measures as well as the acoustic measures F_1, $F_{3dB}^{\rightarrow}$, $\Delta L_{1vsRest}$, S_c, and S_s were considered as dependent variables in separate analyses. [3] We will focus on a discussion of the main and two-way interaction effects, as higher-order interactions are generally difficult to draw conclusions from.

The time-averaged spectral envelopes of performances and their trends across the acoustic descriptors are visualized for bassoon and horn in Figs. 3 and 4, respectively. For the sake of clarity, the data set has been collapsed over the two levels of the 'room' factor, as this factor does not lead to any statistically significant effects. The figures display complete sets of time-averaged spectral envelopes across 32 performances (16 cases $\times$ two rooms) across the eight remaining factor combinations. Furthermore, the corresponding median values for the acoustic descriptors are depicted as well; formant-related descriptors (red) and global descriptors (green). It should be noted that differences in medians computed across participants do not directly correspond to how within-participant variables are evaluated in repeated-measures ANOVAs, with the latter having greater statistical power in detecting effects.

3.1 Main effects

The main effects for 'instrument' are obtained for all acoustic variables, but for none of the behavioral measures. This suggests that the differences are based on systematic deviations between the spectral envelopes of the instruments alone, without bassoonists or hornists judging the assessment of their performances differently. As anticipated and illustrated in Figs. 3 and 4, the spectral-envelope profiles for both instruments bear some resemblance in shape, while notable differences do exist. The strongest differences are found for the descriptors S_c [$F(1, 30) = 36.8$, $p < .001$, $\eta_p^2 = .551$] and F_1 [$F(1, 30) = 21.4$, $p < .001$, $\eta_p^2 = .416$]. While on average the bassoons' main formants are located slightly above 500 Hz, the horns' F_1s lie slightly below that frequency, with an analogous frequency difference for S_c. At the same time, the location of both instruments' $F_{3dB}^{\rightarrow}$ is more similar, reflected in a less pronounced difference [$F(1, 30) = 7.0$, $p = .013$, $\eta_p^2 = .190$]. Differences for the descriptors of relative magnitude differences yield comparable statistical effect sizes (η_p^2).

With regard to within-participant factors, the strongest effects are found for 'voice'. The global descriptor S_c [$F(1, 30) = 165.7$, $p < .001$, $\eta_p^2 = .847$] and the formant descriptors F_1 and $F_{3dB}^{\rightarrow}$ [$F(1, 30) \approx 86.0$, $p < .001$, $\eta_p^2 \approx .740$]

[2] Due to unforeseen technical issues during two experimental sessions, data for a total of five trials were rendered unusable. However, these only concern conditions for which two remaining repetitions were still available, and these were used for the statistical analyses.

[3] Shapiro-Wilk tests on case-based residuals per factor combination yield slight deviations from normality. Across all seven dependent variables and 16 factor combinations, violations are obtained for 23% of tests at $\alpha = .05$, reducing to 6% at $\alpha = .01$. Given the limited number of violations and the known robustness of ANOVAs run on equal sample sizes per factor combination, the statistics are still assumed to be valid.

	bassoon				horn		
vc.	f_0	F_1	$F_{3dB}^{\rightarrow}$	S_c	F_1	$F_{3dB}^{\rightarrow}$	S_c
A	100	100	100	100	100	100	100
B	84	96	97	92	80	89	87
C	63	98	97	91	78	88	86

Table 1. Comparison of frequencies between voice excerpts A, B, and C relative to A (in %), reporting median descriptor values across all non-'voice' factor combinations. The variable fundamental frequency f_0 corresponds to the lowest pitch in each excerpt.

exhibit strong effects, suggesting the presence of systematic differences between unison and non-unison excerpts. As is apparent in the figures, the descriptor frequencies shift downward in the lower-pitched non-unison conditions. Table 1 quantifies these frequency shifts across all performed musical excerpts and compares them to the corresponding shifts in pitch register. Variations in pitch are quantified through f_0 for the lowest pitch occurring in each voice excerpt (i.e., A, B, and C). Although the shifts in descriptor values follow the same trend as for pitch, their deviations remain more constrained compared to the maximum pitch change of 37%. For the bassoon, the formant descriptors are relatively stable and only shift downwards by about 3%, whereas S_c decreases by about 9% for both non-unison excerpts. The horn deviations are most strongly pronounced for F_1, with a downward shift of 21%, whereas the remaining descriptors deviate by about 13%. Across both instruments, $F_{3dB}^{\rightarrow}$ exhibits the weakest dependency on pitch. Overall, these differences appear to stem more from pitch covariation inherent to instrument acoustics than intentional spectral adjustments evoked by the performers.

In addition, the 'voice' factor yields the only main effects with behavioral measures. For the blend ratings, a moderate effect is obtained [$F(1, 30) = 13.3$, $p = .001$, $\eta_p^2 = .308$], which is based on the fact that unison performances lead to higher blend. The weak main effect for musicians' judgments of their performance [$F(1, 30) = 6.0$, $p = .020$, $\eta_p^2 = .168$] is more complex in nature, as it involves several interaction effects and therefore will be discussed further in Section 3.2.

The 'role' factor yields main effects across all acoustic measures. The strongest effects are again obtained for S_c [$F(1, 30) = 95.5$, $p < .001$, $\eta_p^2 = .761$] and $F_{3dB}^{\rightarrow}$ [$F(1, 30) = 31.4$, $p < .001$, $\eta_p^2 = .512$], which yield lower frequencies in the follower condition. This trend is clearly observable in Figs. 3 and 4, especially in the unison conditions that do not exhibit the confounding covariation with pitch discussed above. For the unison conditions in Fig. 3, the relationship between $F_{3dB}^{\rightarrow}$ and S_c is characterized by the latter decreasing relative to the former. Given the stable main formant in this example, the downward shift in centroid implies a reduction of spectral magnitudes located above $F_{3dB}^{\rightarrow}$. Relating these observations to the interaction between performers as a function of their role assignments, followers adjust their spectral envelopes towards being slightly 'darker' in timbre than those of the leaders, without affecting the main formant as much. Along these lines, a single, weak main effect for the factor 'communication' with $F_{3dB}^{\rightarrow}$ [$F(1, 30) = 4.5$, $p = .041$, $\eta_p^2 = .131$] provides another interesting insight. This effect suggests that in the one-way-communication scenario, both musicians perform more 'timidly' by exhibiting lower $F_{3dB}^{\rightarrow}$. In this scenario the leader is unable to hear the follower, which implies that leaders tend to adjust their sounds toward 'darker' timbres, in order to ensure the achievement of blend under the communication impairment.

3.2 Interaction effects

There are several cases of the between-participants factor 'instrument' interacting with the within-participant factors 'role' or 'voice', which are mainly related to effects being more pronounced for the horn, likely due its greater timbral versatility. For example, the horn exhibits more drastic differences along all acoustic measures as a function of performer role as well as being more prone to pitch covariation across different voice excerpts. A similar case concerns the descriptor S_c and an interaction effect 'role' $\times$ 'voice', which is explained by the augmented pitch separation for non-unison voices inducing increased S_c differences between performer roles. In the interest of brevity, no detailed report of the statistics will be made.

As mentioned above, the behavioral measure of individual performance judgments and the 'voice' factor yield complex dependencies based on two-way interactions with 'role' [$F(1, 30) = 6.6$, $p = .015$, $\eta_p^2 = .181$] and 'communication' [$F(1, 30) = 9.5$, $p = .004$, $\eta_p^2 = .241$]. Assuming that the larger effect size conveys the more dominant influence, only in unison performances do musicians rate their performances higher for unimpaired, two-way communication, whereas the ratings for non-unison performances appear to be unaffected by communication directivity. The second interaction involves musicians rating themselves as having performed their role better as followers than as leaders in unison conditions, with the inverse relationship holding for non-unison performances. In addition, the modulating three-way interaction with the additional factor 'instrument' [$F(1, 30) = 4.9$, $p = .035$, $\eta_p^2 = .139$] motivates a reinterpretation with respect to non-unison performances. It suggests that hornists acting as followers rate their performances worse than as leaders, with the contrary applying to bassoonists. This could be related to the playability of the bottom non-unison voice, set in the low pitch register, having been reported as being harder for horns than for bassoons. Overall, these interdependencies suggest that for unison performances, communication impairment has a stronger effect on performers and that followers perform their roles more satisfactorily than leaders.

4. CONCLUSIONS

Both acoustic and behavioral measures succeed in revealing effects of performer interaction within the context of achieving timbre blend. The strongest implication for interaction is found across performer roles. Performers acting as followers adjust their timbres to be 'darker' (i.e., exhibiting lower spectral centroids) compared to their performances as leaders. In the leader role, musicians indicate being less satisfied with their performances, implying that this role bears a larger responsibility for the joint performance (e.g., regarding phrasing, intonation, timing). Hence, leaders may be more critical of their own performance or the resulting blend outcome. In the absence of acoustical feedback from the followers, this increased responsibility may have also encouraged leaders to orient their playing towards avoiding 'brighter' timbres.

Effects found between instruments and between voicings covary with systematic differences in instrument acoustics and pitch range. As a result, the assessment of their actual influence on performer interaction is difficult. This translates to analogous difficulties regarding certain acoustic measures being more sensitive to one instrument or the presence of pitch differences. However, the acoustical analyses based on both pitch-invariant formant traits (e.g., F_1, $F_{\overrightarrow{3dB}}$) and global spectral traits (e.g., S_c) aid in evaluating the different contributions. Across both instruments, $F_{\overrightarrow{3dB}}$ appears least affected by instrument and pitch covariation, and it also leads to the only effect obtained for communication directivity. These observations agree with findings suggesting that $F_{\overrightarrow{3dB}}$ serves as a perceptually salient feature in correlating blend ratings with spectral-envelope traits [10]. Furthermore, the behavioral measures convey that performer interactions appear to be more critical in unison than in non-unison contexts, as the perceived degree of blend is also higher in the first case.

The reported findings will have to be considered preliminary until further analyses are conducted on time-variant datasets. These analyses are expected to provide more insight into effects related to performer interactions that are left concealed in the time-averaged representations as well as allowing two other important influences on timbre blend to be addressed, i.e., intonation and synchrony. While musicians may have succeeded in compensating for effects between room-acoustical environments over the entire duration of performances, the 'room' factor may still become relevant on a finer timescale.

In conclusion, results from this experiment will be valuable to both performance and orchestration practice. For musicians, rules to improve timbre blending between performers could be deduced from effects obtained across musical and acoustical factors. With regard to orchestration, its practitioners will benefit from knowing to what extent performers can affect blend and, conversely, what instrument-specific acoustic properties remain unaffected. These constraints would only emphasize the crucial importance of selecting suitable instrument combinations.

Acknowledgments

The authors would like to thank Harold Kilianski, Yves Méthot, and Julien Boissinot at CIRMMT for their technical assistance. This research was funded in part by a CIRMMT Student Award to Scott Levine and Sven-Amin Lembke as well as a grant from the Canadian Natural Sciences and Engineering Research Council and a Canada Research Chair to Stephen McAdams.

5. REFERENCES

[1] G. J. Sandell, "Roles for spectral centroid and other factors in determining "blended" instrument pairings in orchestration," *Music Perception*, vol. 13, pp. 209–246, 1995.

[2] C. Reuter, *Die auditive Diskrimination von Orchesterinstrumenten - Verschmelzung und Heraushörbarkeit von Instrumentalklangfarben im Ensemblespiel*. Frankfurt am Main: P. Lang, 1996.

[3] W. Goebl and C. Palmer, "Synchronization of timing and motion among performing musicians," *Music Perception*, vol. 26, no. 5, pp. 427–438, 2009.

[4] N. Rimsky-Korsakov, *Principles of orchestration*, M. Steinberg, Ed. New York: Dover Publications, 1964.

[5] C. Koechlin, *Traité de l'orchestration : en quatre volumes*. Paris: M. Eschig, 1954.

[6] G. J. Sandell, "Concurrent timbres in orchestration: a perceptual study of factors determining blend," PhD thesis, Northwestern University, 1991.

[7] C. Reuter, *Klangfarbe und Instrumentation: Geschichte—Ursachen—Wirkung*. Frankfurt am Main: P. Lang, 2002.

[8] K. E. Schumann, "Physik der Klangfarben," Professorial dissertation, Universität Berlin, 1929.

[9] D. Luce and J. Clark, "Physical correlates of brass-instrument tones," *Journal of the Acoustical Society of America*, vol. 42, no. 6, pp. 1232–1243, 1967.

[10] S.-A. Lembke and S. McAdams, "Spectral-envelope characteristics and perceptual timbre blending," In preparation.

[11] ——, "Timbre blending of wind instruments : acoustics and perception," in *Proc. 5th International Conference of Students of Systematic Musicology / SysMus12*, Montreal, Canada, 2012, pp. 1–5.

[12] F. Villavicencio, A. Röbel, and X. Rodet, "Improving LPC spectral envelope extraction of voiced speech by True-Envelope estimation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2006, pp. I–869–I–872.

[13] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The Timbre Toolbox: extracting audio descriptors from musical signals," *Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. 2902–2916, 2011.